Supplementary Information:

Molecular Mechanism of Photoactivation and Structural Location of the Orange Carotenoid Protein in Cyanobacteria

Hao Zhang, Haijun Liu, Dariusz M. Niedzwiedzki, Mindy Prado,Jing Jiang, Michael L. Gross, Robert E. Blankenship

Experimental Method:

Trypsin digestion: The proteolytic digestion protocol was adopted from previous publication with minor changes [1]. Cross-linked sample pellet (~ 0.5 mg protein in total) was re-suspended in 8 M urea (Sigma-Aldrich, St. Louis, MO). Protein was first digested with Lysyl enodpeptidase (Wako Chemical USA, Richmond, VA) at 37°C for 2 hours (enzyme substrate ratio, 1: 50). The sample was further diluted to 1 M urea by 50 mM ammonium acetate (Sigma-Aldrich, St. Louis, MO) solution. The diluted sample was digested with trypsin (Promega, Madison, WI) at 37°C overnight (enzyme substrate ratio, 1:50). Trypsin digestion was quenched by adding trifluoroacetic acid (Sigma-Aldrich, St. Louis, MO).

LC-MS experiment: The peptide solution was further diluted (5 to 10 fold) and analyzed by our LC-MS workflow. We adopted previous setup with some adjustments [2]. A peptide sample was loaded onto an Ultimate 3000 Nano LC system (Thermo Scientific Dionex, Sunnyvale, CA) coupled with LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The peptide was trapped by a guard column (Acclaim PepMap100, 100 μm x 2 cm, C18, 5 μm, 100Å, Thermo Scientific Dionex) by loading the pump with solvent A (water with 0.1 % formic acid, Sigma-Aldrich) at flow rate 6 μl/min. The peptide mixture was separated by a home-packed Michrom Magic C 18 RP column. The peptide was eluted at a flow rate of 300 nl/min ramping a gradient from 5 % to 60 % of solvent B (80 % acetonitrile, 20 % water and 0.1 % formic acid) in 160 min.

The eluted sample was directly introduced into mass spectrometer by the PicoView nano electrospray source (New objective, Woburn, MA). Ion source and parameters of the mass spectrometer were set to spray voltage 1.9 kV, capillary temperature, 200 °C, capillary voltage 80 V and tube lens voltage 100 V. The parameters were optimized by tuning with peptide standards. The mass spectrometer was operated in data-dependent mode with previously reported parameters [1].

Data processing and identification of cross-linked peptides: LC-MS data in Thermo Xcalibur .raw files were converted into mzXML and mgf format by MM File Conversion from Mass Matrix package [3]. Tandem mass spectra were searched against the UniProt protein database by using two search algorithms; Mascot and Mass Matrix. The database searching was refined by using sub-protein database for *Synechocystis* sp. PCC 6803. The protein database for cross-linked peptide search was established from the protein ID list of Mascot searching results[4].

The cross-linked peptides were identified by using two search algorithms: Mass Matrix and *xQuest* [5]. For Mass Matrix, each protein sequence pair was established and search against all LC-MS data. The Mass Matrix search parameters were set as follow: Variable modification: Oxidation of Met, Max # variable PTM/ peptide: 1, Peptide tolerance: 15 ppm, MS/MS tolerance; 0.8 Da, Mass type: Monoisotopic, ^{13}C isotope ions: Yes, Enzyme: Trypsin, Missed cleavages: 3, Fixed modification: none, Peptide length: from 3 to 50, Cross link search mode: Exploratory, Cross link sites cleavability: Non-cleavable by enzyme, Max # cross links/ peptide: 2. The search results was viewed by XMAP (v 0.5.1, Mass Matrix). For *xQuest*, search parameters were set as previous report[1, 5].

All positive hits for cross-linked peptides from both search algorithms were manually validated. The MS spectra of cross-linked peptide were manually inspected. For tandem MS spectra, the sequence of ions without covering cross linked lysine residues were checked separately (Fig. S6).

Figure S1

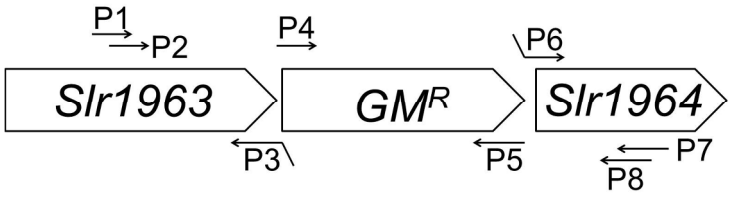


Figure S1. The Construction of OCP-His strain. The gene *Slr1963* locus was genetically modified and C-terminally His_6-tagged OCP was expressed. A gentamicin resistant gene cassette is located between the *Slr1963* and *Slr1964* genes. Positions of primers used in the construction and segregation analysis are labeled and primer sequences are listed in Table S1.

Figure S2

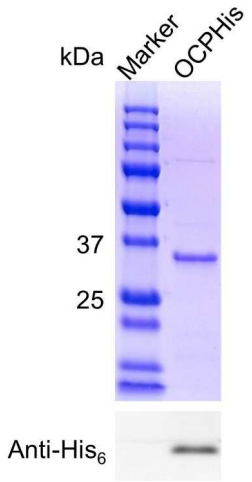
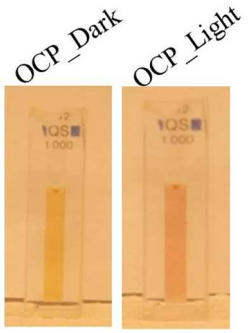


Figure S2. SDS-PAGE (upper panel) analysis of purified C-terminally His_6-tagged OCP. Immunodetection (lower panel) of the polyhistidine tag of the genetically modified OCP.

Figure S3

(A)



(B)

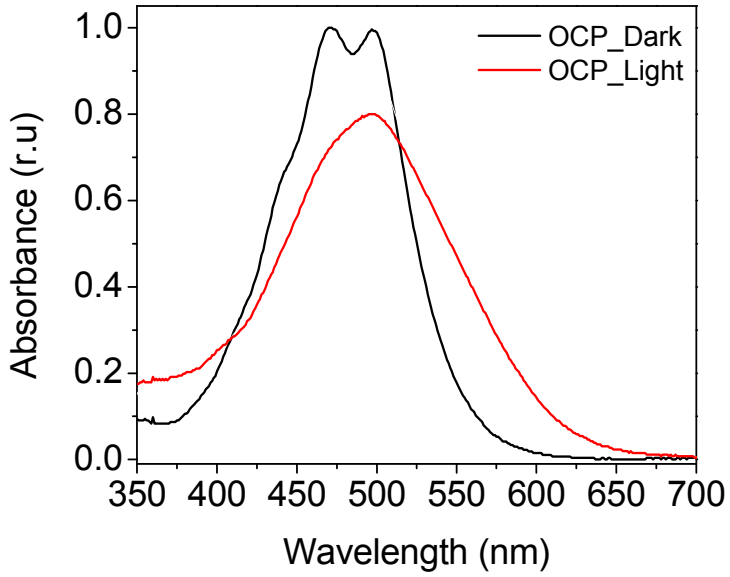
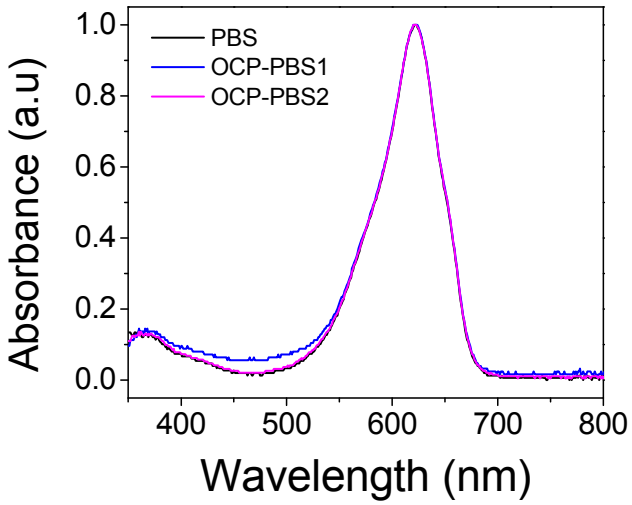


Figure S3. **Light-induced absorption spectrum change of the OCP.** (A) Pictures of isolated dark adapted OCP (OCP_Dark) and light illuminated OCP (OCP_Light). White light with 1500 µmol photons $m^{-2} \cdot s^{-1}$ intesity was used for 5 min to initiate and maintain photoactivation. (B) Absorption spectra of the inactive "orange" form (OCP_Dark; black) and the light activated "red" form (OCP_Light; red).

Figure S4

(A)



(B)

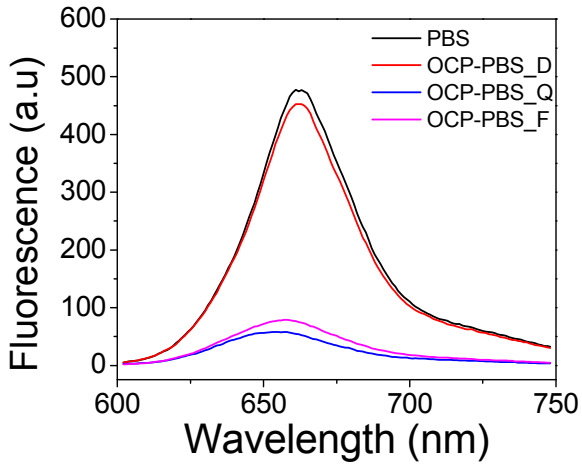
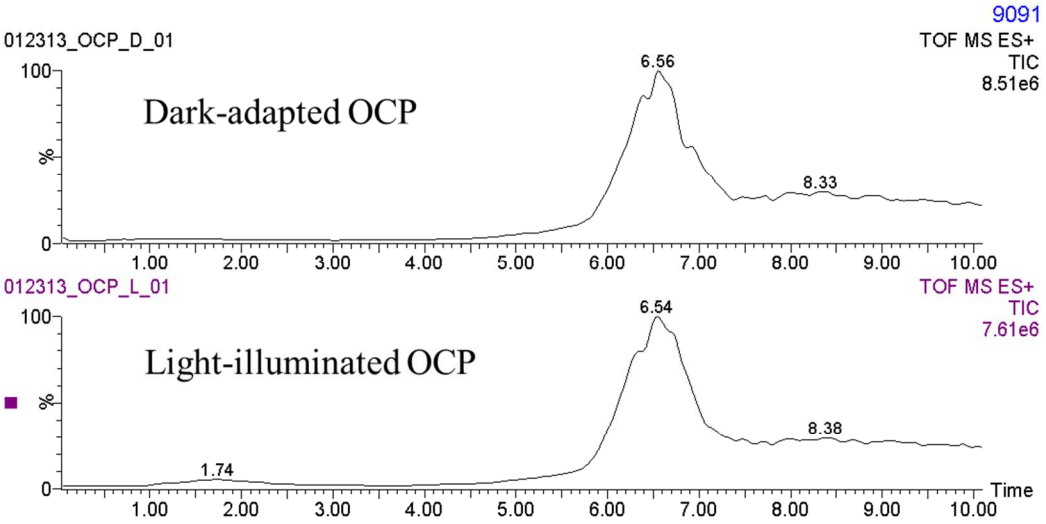


Figure S4. The OCP-PB reconstitution and fluorescence quenching of PB. (A) The room temperature absorption spectra of PB (black line), OCP-PB mixture with an excess of OCP (OCP-PB1; blue line), absorption of OCP-PB complexes after removal of excess OCP by filtration (magenta line). (B) The room temperature fluorescence emission spectrum of dark-adapted PB (black line), OCP-PB mixture unquenched by OCP (OCP-PB_D; red line), OCP-PB mixture quenched by OCP (OCP-PB_Q; blue line), OCP-PB complex after removal of excess of the OCP, PB fluorescence quenched with stoichiometric binding of OCP to PB (OCP-PB_F; magenta).

Figure S5

(A)



(B)

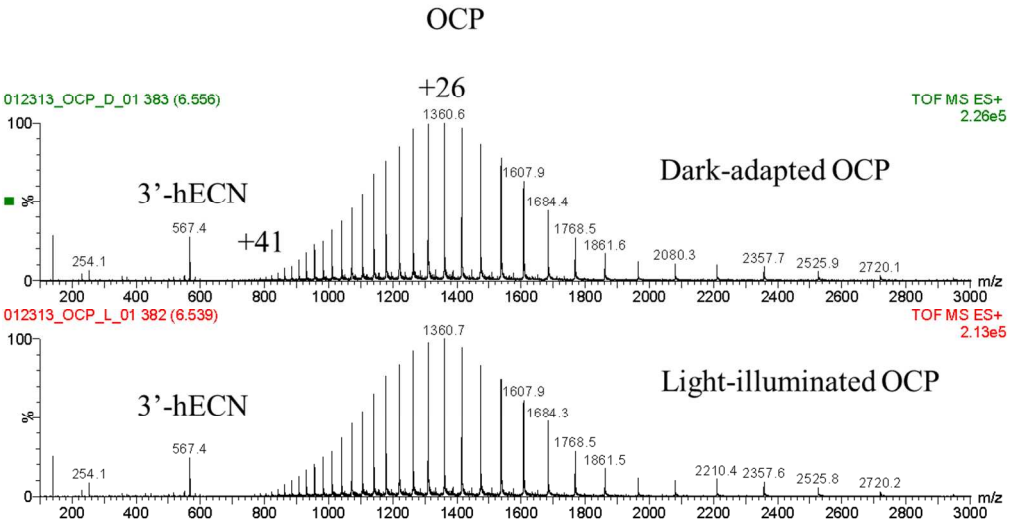
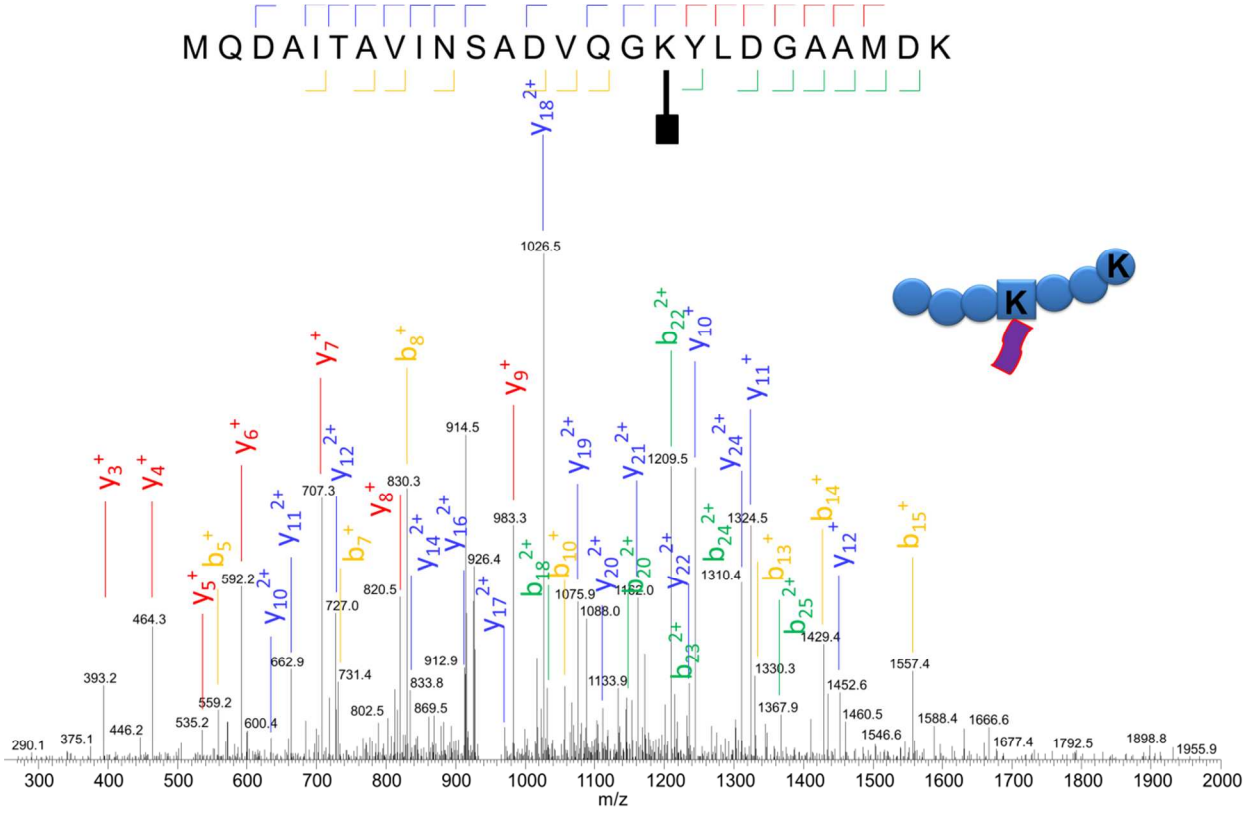


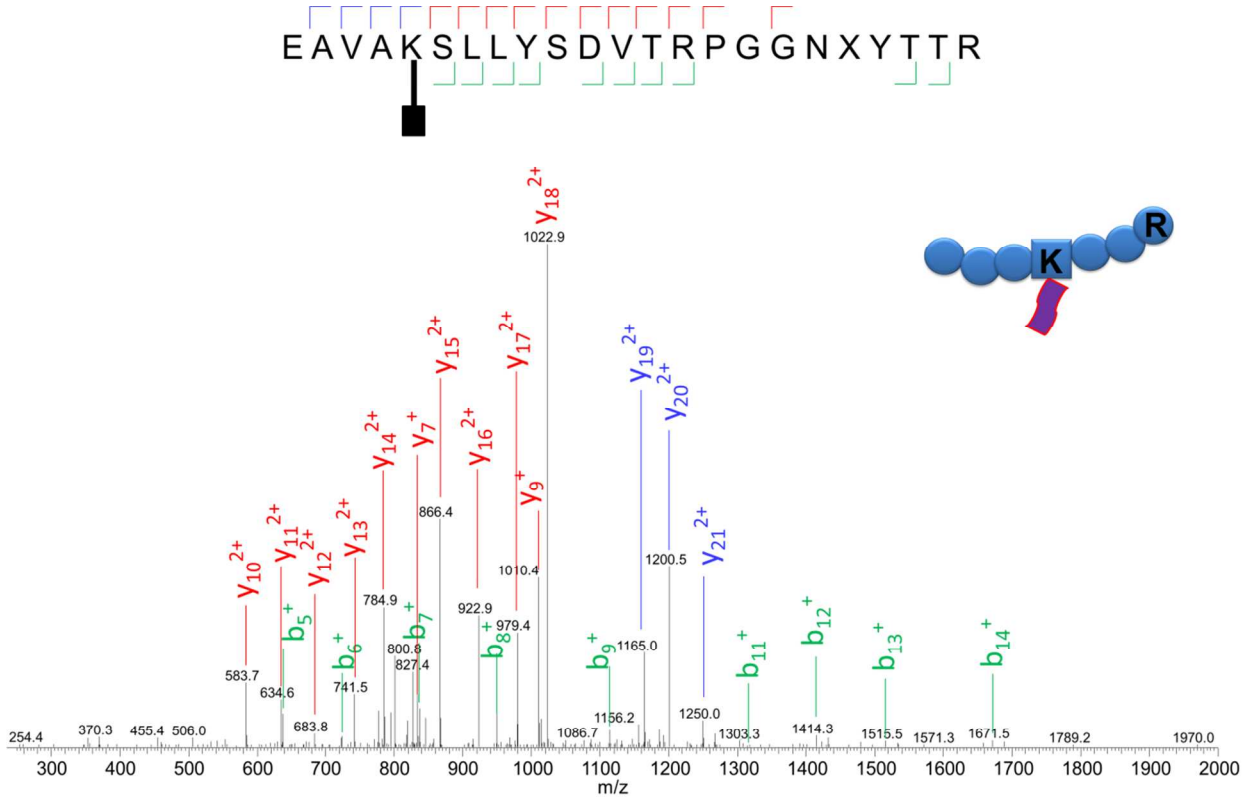
Figure S5. The regular LC-MS results of intact OCP. (A) The LC chromatograms of dark-adapted and light-illuminated OCP. The OCP protein was eluted in the same time (center at 6.5 min) during the reverse phase HPLC. There is no other protein signal detected in the isolated OCP sample. (B) The mass spectra of dark-adapted and light-illuminated OCP. The pigment, 3'-hECN, was released and observed at m/z 567.4. The multiple-charged OCP (up to 41) peaks were observed in both samples. The molecular weight (MW) measured in both samples are same.

Figure S6

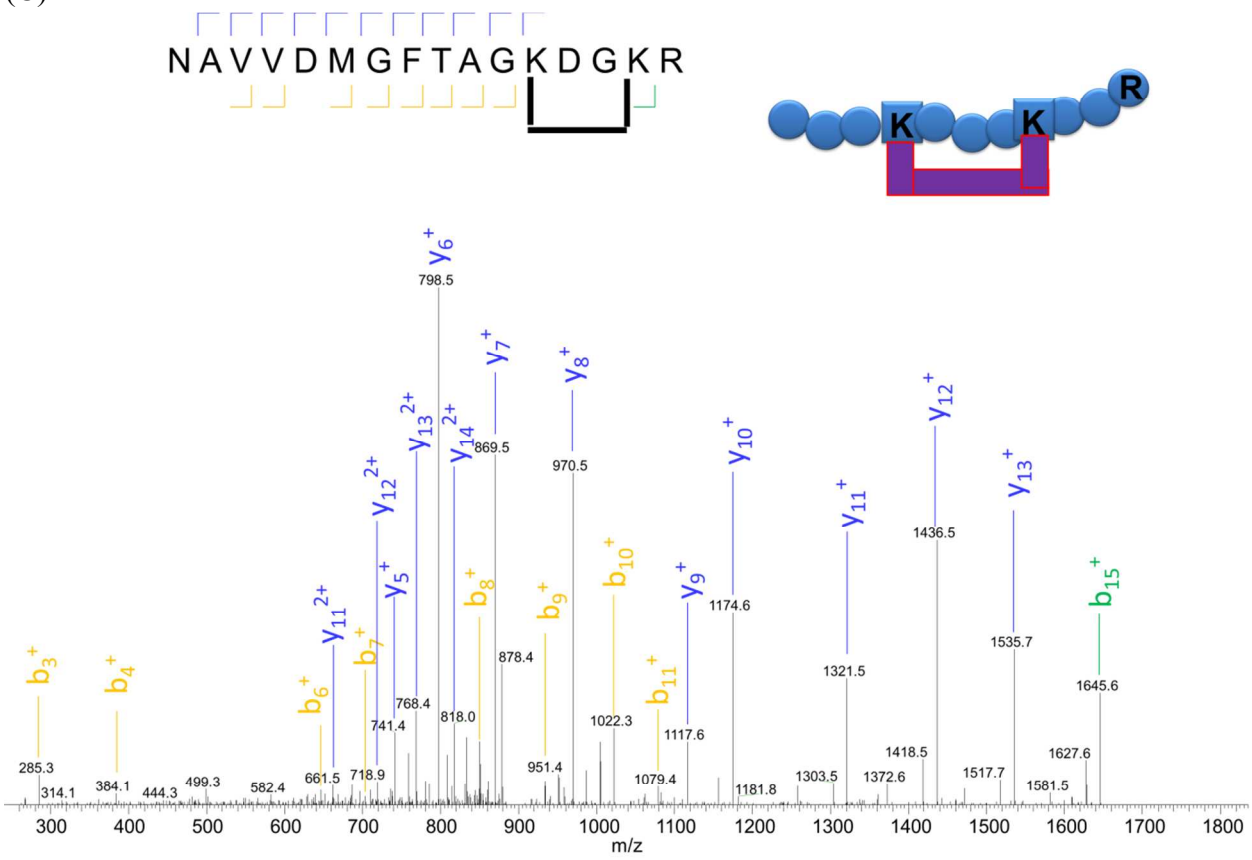
(A)



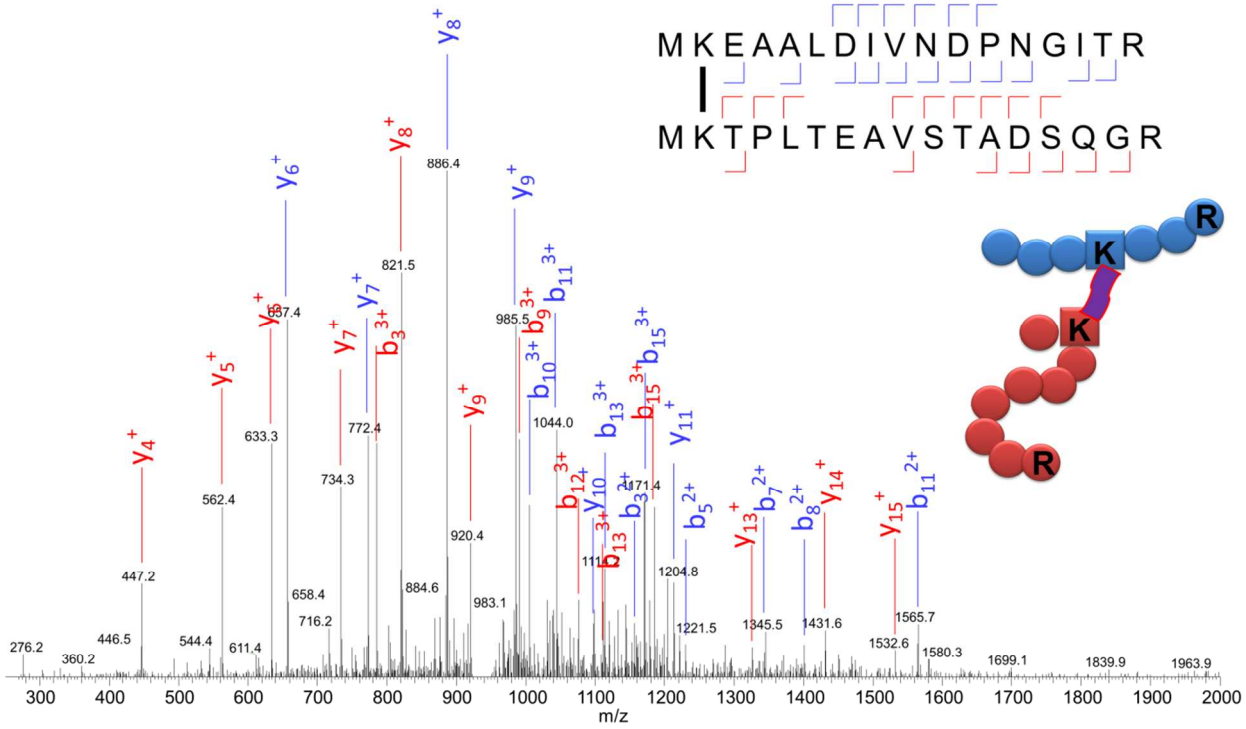
(B)


EAVAKSLLYSDVTRPGGNXYTTR
K
R
m/z

(C)



(D)



(E)

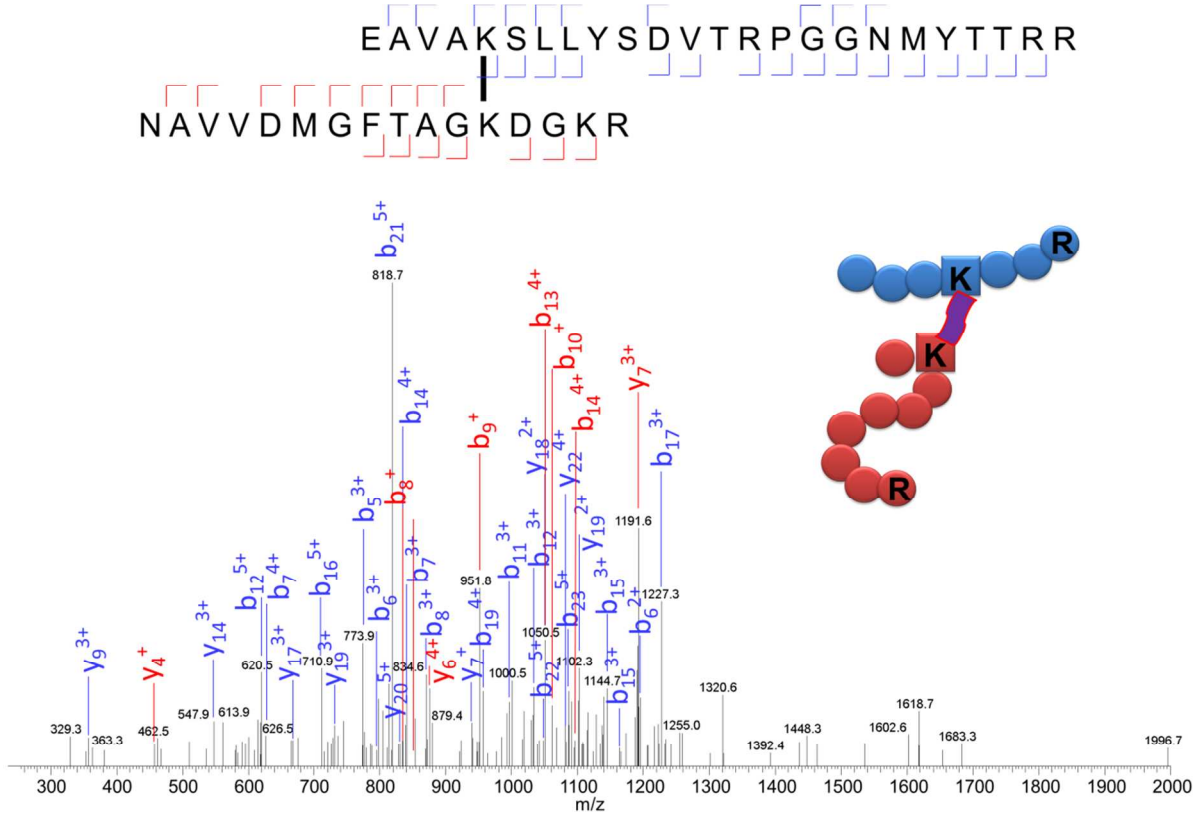


Figure S6. LC-MS/MS spectra of cross-link peptides. (A) Tandem MS spectrum of mono-linked peptide, peptide 1-26 from ApcB subunit. Peaks of sequence b and y ions are labeled in the spectrum. Peaks of sequence b ions that do not contain dead-end cross links are labeled in orange. Peaks of sequence b ions that contain dead-end cross links are labeled in green. Peaks of sequence y ion are labeled by two colors: red (peaks without dead-end cross links) and blue (peaks with dead-end cross links). The modified 17K (K-lysine) residue is located. (B) Tandem MS spectrum of mono-linked peptide, peptide 54-76 from ApcB subunit. The modified 58K residue is located. (C) Tandem MS spectrum of loop-link peptide, peptide 156-171 from OCP subunit. Peaks of sequence b and y ions are labeled in the same way as the mono-linked peptide. Two lysine residues are linked, 167K to 170K, within one peptide. (D) Tandem MS spectrum of inter-linked peptide, peptide 1-17 from CpcA and peptide 134-150 from CpcB. Peaks of sequence b and y ions from peptide 1-17 (CpcA) are labeled in red. Sequence b and y ions from peptide 134-150 (CpcB) are labeled in blue. 2K (CpcA) and 135K (CpcB) are the linked sites in the inter-linked peptide. In three tandem MS spectra, most of the major peaks are labeled. Because the same sequence ion can have different peaks in the tandem MS spectra, like different charge states, neutral loss of ammonia or water, several peaks from sequence b and y ions with loss of ammonia or water are not labeled in the tandem spectrum. (E) Tandem MS spectrum of inter-linked peptide, peptide 54-76 from ApcB and peptide 156-171 from OCP. Peaks of sequence b and y ions from peptide 54-76 (ApcB) are labeled in red. Sequence b and y ions from peptide 156-171 (OCP) are labeled in blue. 58K (ApcB) and 167K (OCP) are the linked sites in

the inter-linked peptide. In three tandem MS spectra, most of the major peaks are labeled. Because the same sequence ion can have different peaks in the tandem MS spectra, like different charge states, neutral loss of ammonia or water, several peaks from sequence b and y ions with loss of ammonia or water are not labeled in the tandem spectrum.

Figure S7

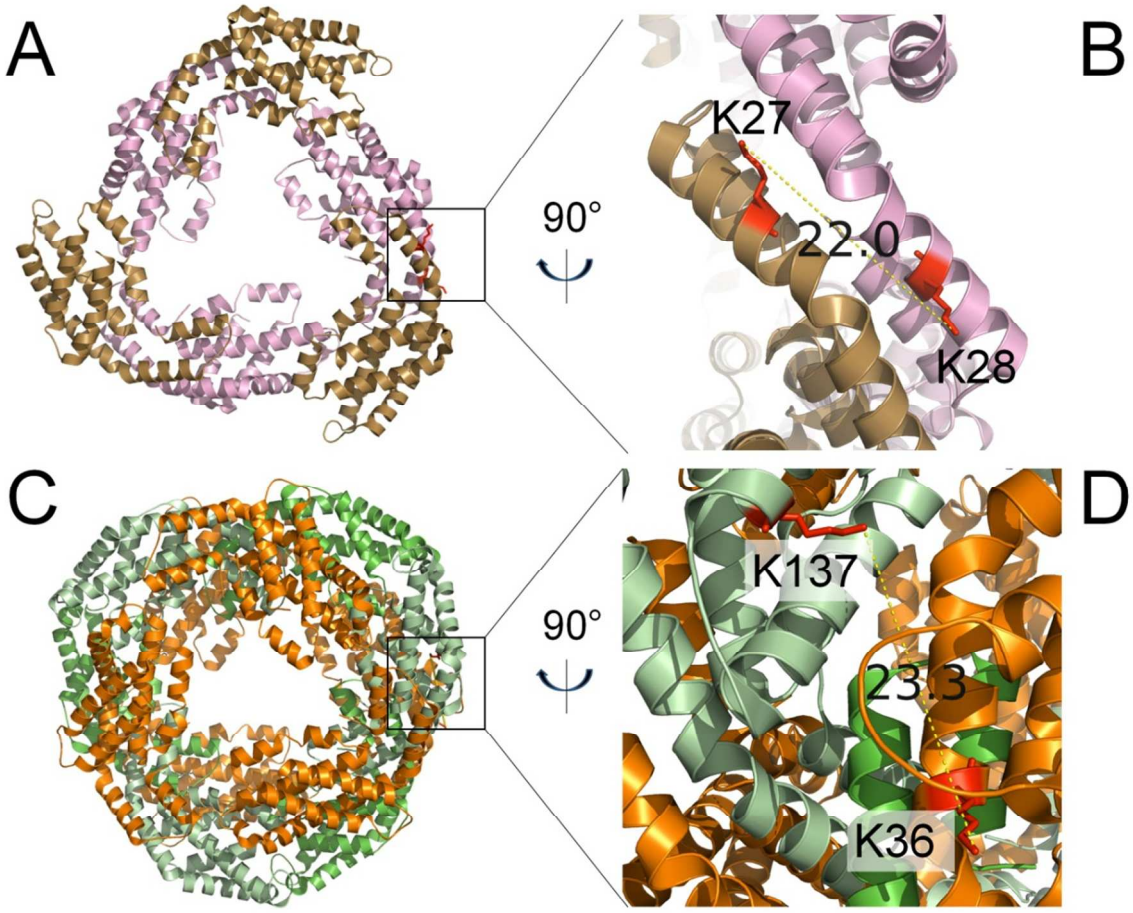


Figure S7. (A, B) Cross-links identified in Apc trimer (PDB ID: 4F0U) and (C, D) Cpc trimer of trimer (PDB ID: 4H0M). LC-MS/MS indicated ApcA:K27-ApcB:K28 and CpcA:K137-CpcB:K36 cross-links. ApcA (cartoon, sand), ApcB (cartoon, pink), CpcA (pale green, green from two trimers) and CpcB (orange) (Table S3). All lysine residues are labeled as red sticks. The distances between cross-linked species are indicated (Å).

Tabel S1: Primer used in the construction of Figure S1

Primer	Sequence
OCP1F(P1)	GGCTTCTGGTACCGTTTAG
OCP2F(P2)	CTGATGGAGCAGGGCTTTG
OCPGMR(P3)	GTTTTACAACGTCGTGACTGGGAAACTAGTGATGGTGATGGTGATG
GM1F(P4)	CACACCATCACCATCACTAGTTTCCCAGTCACGACGTTGTAAAAC
GM1R(P5)	ACTCATTAGGCACCCCAGGC
OCPdownF(P6)	GTGTAAAGCCTGGGGTGCCTAATGAGTGAATAACTCCCTTCAGAGTTTTGTCTTTGCC
OCPdownR1(P7)	AGGCTCTAGCATTCCTCATC
OCPdownR2(P8)	GGACCGGGAAAATGTATCCTTTGCCG

Table S2: Sequence coverage of the LC-MS/MS identified PB-OCP reconstituted sample.

Protein Subunit	Sequence Coverage	Mass (Average)	# of Peptides Identified	# of Spectral Matches
ApcA	97%	17412	29	222
ApcB	97%	17216	26	129
ApcC	85%	7805	9	22
ApcC2	96%	30797	63	498
ApcD	38%	17923	5	9
ApcE	95%	100295	169	750
ApcF	96%	18892	21	173
CpcA	98%	17586	43	223
CpcB	92%	18126	45	459
CpcC1	93%	32520	60	397
CpcD	89%	9322	14	79
CpcG1	94%	28902	65	277
CpcG2	76%	28522	17	29
Ferredoxin-NADP oxidoreductase	92%	46359	75	144
FRP	82%	15372	16	29
OCP	94%	34659	57	219

Table S3: LC-MS/MS identified cross-links after OCP-PBS reconstitution.

Sequence	Protein A	Protein B	m/z	Charge
IK(27)AFVTGGAAR-LK(28)SYFASGELR-a2-b2	ApcA	ApcB	625.3516	4
EAVAK(58)SLLYSDVTRPGGNXYTTRR-NAVVDXGFTAGK(167)DGKR-a5-b12	ApcB	OCP	904.8656	5
YLDGAAMDK(26)LK-VDK(685)EVTPR-a9-b3	ApcB	ApcE	577.0588	4
K(331)QFFEPFINSR-RLAK(325)SPLYR-a1-b4	ApcE	ApcE	664.1208	4
XSVK(4)ASGGSSLARPQLYQTVPVSAISQAEQQDR-DGK(170)RIAEPVVPPQDTASR-a4-b3	ApcE	OCP	1116.5800	5
GEYLSGSQLDALSATVAEGNK(36)R-YIK(137)ANHGLSGDAR-a21-b3	CpcB	CpcA	761.7911	5
MK(135)EAALDIVNDPNGITR-MK(2)TPLTEAVSTADSQGR-a2-b2	CpcB	CpcA	947.2343	4
QIFFHAFK(59)WDR-EK(64)VLESQLR-a8-b2	CpcG1	CpcG1	684.1190	4
EECQNLK(249)LIPER-YLDGAAXDK(26)LK-a7-b9	OCP	ApcB	713.1163	4

[a]Listed cross-links are representative and used as internal references, see Fig. S5 for examples. Some cross-links between other components of PBS are not listed in this table and will be published separately.

Reference

1. Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F. U., Ban, N., Malmstrom, L., and Aebersold, R. (2012) Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry, *Science 337*, 1348-1352.
2. Zhang, H., Huang, R. Y., Jalili, P. R., Irungu, J. W., Nicol, G. R., Ray, K. B., Rohrs, H. W., and Gross, M. L. (2010) Improved mass spectrometric characterization of protein glycosylation reveals unusual glycosylation of maize-derived bovine trypsin, *Analytical chemistry 82*, 10095-10101.
3. Xu, H., Hsu, P. H., Zhang, L., Tsai, M. D., and Freitas, M. A. (2010) Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry, *Journal of proteome research 9*, 3384-3393.
4. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis 20*, 3551-3567.
5. Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Forster, F., Beck, M., and Aebersold, R. (2012) False discovery rate estimation for cross-linked peptides identified by mass spectrometry, *Nature methods 9*, 901-903.